基于高斯泼溅的轻量级重建场景分割方法

王 锋[1),2)] 银 莹[1),2)] 王佳炎[1),2)] 唐 勇[1),2)] 李 胜[3),4)] 赵 静[1),2)]

[1)](燕山大学信息科学与工程学院 河北 秦皇岛 066004)

[2)](河北省计算机虚拟技术与系统集成重点实验室 河北 秦皇岛 066004)

[3)](北京大学计算机学院 北京 100871)

[4)](北京大学北京市虚拟仿真与可视化工程中心 北京 100871)

摘 要 在三维重建领域，最新提出的三维高斯溅射方法相比于神经辐射场方法在训练时间和重建质量上均得到显著提升，但是由于重建场景的复杂性和不同物体之间产生的遮挡，导致重建场景的分割十分具有挑战性。本文提出了一种轻量级的3D高斯重建场景分割方法，能够在复杂的重建场景中分割出指定物体。具体来说，首先利用高斯泼溅的可微分特性，在重建过程中为每个高斯核训练一个类别特征用于类别预测。同时，设计并训练一个多层感知器作为解码器，输入类别特征预测出高斯核的类别信息，从而实现对目标对象的高效分割。接着，在渲染图像的同时，将类别特征映射到每个像素上，并使用图像分割模型获得目标物体对应区域的类别特征，再将其输入解码器以获得所选物体的类别，从而实现在重建场景中分割目标物体。最终，为了减少分割过程中产生的噪声并优化分割结果，在分割结果中进一步使用KNN算法对分割结果进行去噪处理。最终实现从任意视角分割重建场景中的目标对象。实验结果表明，本文方法能够有效应用于多种复杂场景，可轻松地与现有的图像分割模型集成，实现毫秒级内对重建场景的分割，在节省大量内存的同时不影响分割质量。相关代码将发布在 https://github.com/wangfeng70117/Gaussian-Extracting.

关键词 三维分割；神经辐射场；三维重建；高斯泼溅；分割模型

中图法分类号 TP391 **DOI号** 10.11897/SP.J.1016.2025.01232

Object Segmentation in 3D Reconstructed Scenes Based on Gaussian Splatting

WANG Feng[1),2)] YIN Ying[1),2)] WANG Jia-Yan[1),2)] TANG Yong[1),2)] LI Sheng[3),4)] ZHAO Jing[1),2)]

[1)](School of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei 066004)

[2)](Hebei Key Laboratory of Computer Virtual Technology and System Integration, Qinhuangdao, Hebei 066004)

[3)](School of Computer Science, Peking University, Beijing 100871)

[4)](Beijing Engineering Technology Research Center of Virtual Simulation and Visualization, Peking University, Beijing 100871)

Abstract In the field of 3D reconstruction, the recent introduction of 3D Gaussian splatting has led to substantial advancements in both training efficiency and reconstruction quality, especially when compared to traditional neural radiance field-based approaches. These advancements make 3D Gaussian splatting a promising tool for reconstructing complex scenes with high accuracy and efficiency, significantly reducing the training time compared to the methods based on Neural Radiance Fields. However, challenges remain in effectively segmenting the reconstructed scenes, particularly when dealing with intricate occlusions and overlapping objects. Scene segmentation in

收稿日期：2024-10-19；在线发布日期：2025-03-05。本课题得到河北省省级科技计划资助(中央引导地方：246Z0105G)、河北省创新能力提升计划项目资助(22567626H)。**王 锋**，博士研究生，中国计算机学会(CCF)会员，主要研究方向为计算机图形学、计算机视觉、三维重建、物理仿真。E-mail：wangfeng70117@stumail.ysu.edu.cn。**银 莹**，硕士研究生，主要研究方向为计算机视觉、三维重建。**王佳炎**，博士研究生，中国计算机学会(CCF)会员，主要研究方向为计算机图像处理、图像水印。**唐 勇**，博士，教授，中国计算机学会(CCF)高级会员，主要研究方向为虚拟现实技术及应用。**李 胜**，博士，教授，中国计算机学会(CCF)高级会员，主要研究方向为计算机图形学、虚拟仿真技术。**赵 静(通信作者)**，博士，副教授，中国计算机学会(CCF)会员，主要研究方向为虚拟现实技术及应用、计算机视觉、物理仿真。E-mail：zhaojing@ysu.edu.cn。

such complex environments becomes difficult due to the dynamic interactions between objects, subtle boundaries, and the presence of noise in the reconstruction process. These challenges hinder the accurate identification and isolation of specific objects within the scene, which is critical for a range of practical applications. To address this issue, this paper proposes a novel, lightweight method for 3D Gaussian-based scene segmentation that effectively segments specific objects from a reconstructed scene. The core innovation of this approach is the use of the differentiable nature of 3D Gaussian splatting during the reconstruction process, which allows for the simultaneous training of object class features alongside the scene reconstruction. Specifically, a class feature is associated with each Gaussian kernel, encoding semantic information that is crucial for object recognition and classification within the 3D space. These class features are trained to distinguish between different objects and map them to specific object categories. After embedding the class features into the scene, a multi-layer perceptron decoder is designed and trained to map the class features to object class labels, allowing the system to identify and classify objects based on the features associated with each Gaussian kernel. During the rendering phase, the class features are mapped onto 2D image pixels, where an image segmentation model is employed to obtain the corresponding class features for the target object's region. This segmentation model helps isolate the desired object from the entire reconstructed scene, enabling the decoder to predict the class of the object with high accuracy. Additionally, to further enhance segmentation quality and reduce noise, the output of the segmentation is refined using the k-nearest neighbor algorithm, which helps eliminate false positives and smooths the boundaries of the segmented object. This post-processing step ensures that the segmented object is accurately isolated, and the method can be applied to any viewpoint within the reconstructed scene. Experimental results demonstrate that this approach performs well on a variety of complex scenes, achieving real-time segmentation in milliseconds and seamlessly integrating with existing image segmentation models. Furthermore, the method offers significant memory savings as it avoids the large-scale memory usage associated with traditional segmentation techniques, making it not only highly efficient but also scalable. This makes it highly efficient and scalable, with the potential for real-time applications in interactive environments. Overall, the proposed 3D Gaussian-based segmentation method represents a promising solution to the challenging problem of object segmentation in highly complex 3D reconstructed scene, offering a significant step forward in the field of 3D scene reconstruction and enabling more realistic simulations and interactive applications. The code will be released at https://github.com/wangfeng70117/Gaussian-Extracting.

Keywords　3D segmentation; neural radiance field; 3D reconstruction; Gaussian splatting; segmentation model

1　引　言

3D 场景理解主要涵盖 3D 场景重建和 3D 场景感知，是计算机图形学与计算机视觉领域的一个富有挑战性的任务，在场景操控和虚拟现实等应用中具有重要意义。许多研究探索了多种 3D 表示形式以及 3D 分割方法，包括 RGB-D 图像[1-2]、点云[3-4]、体素[5-6]和鸟瞰视图空间[7-8]。相较于 2D 分割技术，3D 分割仍处于早期发展阶段，数据标注的稀缺性与高计算复杂度使得类似 SAM[9]及其变体[10]的统一框架面临较大挑战。为应对数据匮乏的问题，已有多项研究[11-12]探索了从 2D 基础模型向 3D 模型迁移的方法。

近年来，神经辐射场(Neural Radiance Fields，NeRF)[13]等隐式 3D 表示方法在 3D 场景重建领域取得了显著进展，3D 检测和语义分割等 3D 场景感知任务也在不断深入研究。尽管大量实验表明，现

有的基于 NeRF 的方法在 3D 场景理解中取得了显著成功，但由于 NeRF 的高计算耗时和 3D 数据采集成本，这些方法的扩展受到很大限制。此外，NeRF 的训练和渲染过程所需的大量时间使其难以应用于交互操作，从而限制了其在实际任务中的应用。而作为近年来三维重建领域中的一种热门技术，三维高斯泼溅(3D Gaussian Splatting，3DGS)[14]具有更高的重建精度和实时渲染能力，显著地降低了时间成本。3DGS 通过将密集点云信息训练为一系列三维高斯分布，并在渲染管线中将这些高斯分布的颜色信息映射到图像上，计算每个像素的颜色，从而生成渲染图像。这种显式表示方法明显地提升了渲染效率，能够实现实时渲染。因此，基于 3DGS 的方法被广泛选用于三维场景感知领域中的目标分割任务。

由于 3DGS 具备高质量和实时渲染能力，许多研究人员采用显式表示来刻画复杂物体的形状和外观，并不断探索其在更多领域的应用。尽管近期提出了多种基于 3DGS 的场景分割方法，但是这些方法仍存在一些局限性。例如，Gaussian Grouping[15]实现了场景分割和物体移除任务，但其内存需求过大，难以在普通显卡上进行训练。SAGA[16]的实现过程较为复杂，此外，由于每个高斯分布可能同时对应多个物体，准确分割单个物体依然具有挑战性。Feature-3DGS[17]需要大量的训练时间，并且仅支持文本分割，文本提示需要属于训练时的标签数据集，无法实现开放性语义查询。虽然相比于基于 NeRF 的三维场景表示方法，3DGS 在重建质量和速度方面取得了显著提升，但在三维场景分割领域，3DGS 仍然面临若干挑战，尤其是在应用的灵活性和实现复杂性方面。

为进一步支持更简洁高效的模型分割，本文提出了一种基于 3DGS 的轻量级三维场景分割方法，该方法能够从任意视角实现精确的三维场景分割。具体而言，在训练三维场景的过程中，通过 3DGS 的可微分渲染框架为每个高斯分布训练了一个长度为 32 的类别特征，同时训练了一个多层感知器(Multilayer Perceptron，MLP)作为解码器用于对高斯核进行类别预测。为了实现场景分割，在交互选择场景中的物体后，首先从当前视角渲染图像，并利用图像分割模型对选中的物体进行分割，然后将分割图像上的特征和高斯核上的特征输入到解码器进行解码，从而分割出对应高斯核。随后，使用 K 近邻算法(K-Nearest Neighbor，KNN)从提取的高斯核中去除噪声点，以增强模型的细节表现。最终，实现了基于 3DGS 的模型精准分割，能够快速高效地从三维场景中提取指定物体。

本文的主要贡献总结如下：

(1)提出了一种高效的高斯分割方法，能够实现更加准确的分割结果，同时降低训练内存开销。

(2)本方法支持从任意视角分割 3DGS 重建模型，并且能够轻松地与现有 2D 分割模型相结合。

(3)本方法有效地解决了由于场景复杂性导致的模型边界模糊和噪声问题，从而提高了分割结果的质量和准确性。

2 相关工作

2.1 三维重建

在三维重建领域，研究人员提出了多种三维表示方法。NeRF 作为一种采用体渲染技术并利用二维数据进行监督学习来训练三维模型的开创性工作，被视为该领域的一个重要里程碑。进而，众多研究进一步提升了 NeRF 的重建质量。例如，Mip-NeRF[18]通过高效渲染抗锯齿的圆锥截锥体，有效减少了锯齿伪影，并显著提升了 NeRF 在细节表现上的能力。Mip-NeRF-360[19]进一步扩展了 Mip-NeRF，采用非线性场景参数化、在线蒸馏和基于畸变的正则化方法，成功解决了无界场景的挑战。Zip-NeRF[20]则结合了 Mip-NeRF-360 与基础的网格模型，降低了重建错误率。MVSNeRF[21]和 PixelNeRF[22]基于不同视角获取的图像特征构建了通用的三维表示。

尽管 NeRF 在三维重建领域取得了突破性进展，但其训练和渲染效率相对较低，并且在训练过程中需要占用大量内存。为了解决这些问题，后续研究逐渐转向将显式特征与隐式辐射场相结合，以提高性能。最近，3DGS 凭借其高质量的渲染效果和实时渲染能力脱颖而出，逐渐取代了 NeRF 成为三维表示的主流方法。与 NeRF 的体渲染不同，3DGS 采用一系列三维高斯核来表示场景中的物体。每个高斯核包含均值、协方差矩阵、不透明度以及球谐函数等参数。通过基于点的渲染方法，3DGS 高效地实现了从三维到二维的投影，并支持实时渲染。目前，3DGS 技术已广泛应用于多个研究领域，并取得了显著进展。例如，动态 3D 高斯[23]利用高斯分布跟踪动态场景中的三维物体，将 3DGS 扩展到动态场景的表示与渲染。同样，4D

高斯泼溅[24]进一步发展了这一方法，能够有效地处理复杂的动态场景。除此之外，在模型生成领域，DreamGaussian[25] 和 GaussianDreamer[26] 将3DGS和扩散模型[27]相结合，实现了使用文本提示生成高质量三维模型的方法。

通过采用这些隐式、显式和混合的三维表示方法，研究人员能够更准确地重建三维环境，完成从三维场景重建到感知的多个方面的任务。在这一背景下，三维场景分割成为一种功能多样且具有探索潜力的方法，进一步推动了该领域的发展。

2.2　神经辐射场分割

三维场景分割是计算机视觉和计算机图形学中的一个基础且极具挑战性的问题。传统的三维资产通常以点云或网格形式表示，学术界已经提出了许多基于深度学习的三维分割方法。随着 NeRF 在新视角合成方面的成功，越来越多的研究者开始探索使用 NeRF 进行三维物体分割，并将其应用于语义分割和三维重建等任务。例如，Semantic-NeRF[28]将语义信息融入外观和几何结构中，展示了 NeRF 在标签传播和精化方面的潜力。NVOS[29] 和 SA3D[30]通过语义标签、用户输入和二维掩码等提示，指导三维场景的分割。LERF[31] 和 SRF[32] 利用二维自监督模型，训练与 NeRF 对齐的附加特征场，旨在将二维视觉特征提升至三维空间中。DM-NeRF[33]引入了对象场，通过生成空间中每个点的对象独编码实现分割。SPIn-NeRF[34]通过语义辐射场来评估场景中某个位置与特定物体的相关性。此外，OR-NeRF[35]通过将二维分割结果投影回三维空间中，减轻了计算负担，同时确保三维一致性。该方法在重新渲染二维图像之前，还引入了深度监督和感知损失。

尽管这些方法展现了在三维场景分割中的潜力，但它们仍然受到 NeRF 渲染效率的限制。在如何高效地执行三维场景分割，并将其应用于场景编辑、场景理解和碰撞检测等任务时，仍然面临诸多挑战。

2.3　高斯泼溅分割

凭借在复杂场景表示和训练时间的卓越表现，3DGS 已广泛应用于三维场景理解和编辑。近年来，许多研究者专注于 3DGS 中的分割领域，并取得了显著进展。Gaussian Grouping[15]结合 SAM[9]生成的二维掩码，通过三维空间一致性正则化来监督可微渲染过程中的类别代码。这些代码用于指导三维场景中物体的高斯核分组，同时开发出了一种局部高斯核的编辑方法，但是该方法存在严重存消耗问题。Feature-3DGS[17] 和 LangSplat[36] 从 CLIP[37]的语言特征编码到每个高斯核上，创建了语言场，以便通过文本提示实现高斯核的分割，但交互性较为有限，并且需要大量的训练时间。SAGA[16]则结合了二维分割基础模型与 3DGS，通过将多尺度的二维分割结果嵌入到 3D 高斯核的特征中，实现了毫秒级的三维分割。然而，在这一过程中，每个高斯核可能对应多个物体，导致分割结果不准确。

尽管这些基于 3DGS 的分割方法避免了 NeRF 固有的消耗问题，但是仍面临其他挑战。例如训练内存开销过大，实现方法过于复杂，交互方式比较单一等，难以广泛应用于实际场景。针对这些挑战，本文提出了一种轻量级的分割方法，能够从重建的三维场景中快速，精确地从任意角度分割指定物体。同时，该方法旨在增强分割的交互性，并在训练过程中最大程度地减少内存开销，以提高其实际应用价值。

3　本文方法

本文设计了一种轻量级且高效的重建场景目标分割方法，整体管线如图 1 所示。

本文方法首先在每个 3D 高斯核上训练长度为 32 的类别特征，以保留更多的类别信息。在分割时，将分割出物体区域的类别特征输入到解码器中进行类别预测获得目标物体的类别，同理，将高斯核上的类别特征也输入到解码器获得每个高斯核的类别，找到对应的高斯核从而实现目标物体的分割。最后，使用去噪算法解决分割目标不准确问题。本文方法能和已有的二维分割模型集成，从重建场景的任意视角分割出指定物体。

3.1　三维高斯泼溅

作为三维重建领域的最新进展，3DGS 提出了一种高精度的可微光栅化算法。3DGS 使用一组图像数据集 I 及其相应的相机位姿训练一组高斯核 $\mathcal{G}$ 以表示 3D 场景，每个高斯核包含位置 μ 、协方差矩阵 $\boldsymbol{\Sigma}$ 、不透明度值 α 以及用于表示颜色的球谐系数 SH 。对于特定的相机位姿，3DGS 首先将高斯核投影到渲染图像上，并通过 α 混合的方式计算投影高斯核对每个像素的影响权重和对覆盖像素的颜色贡献，其权重的计算方式为

$$T(x)=e^{-\frac{1}{2}(x-\mu)^{T}\boldsymbol{\Sigma}^{-1}(x-\mu)} \tag{1}$$

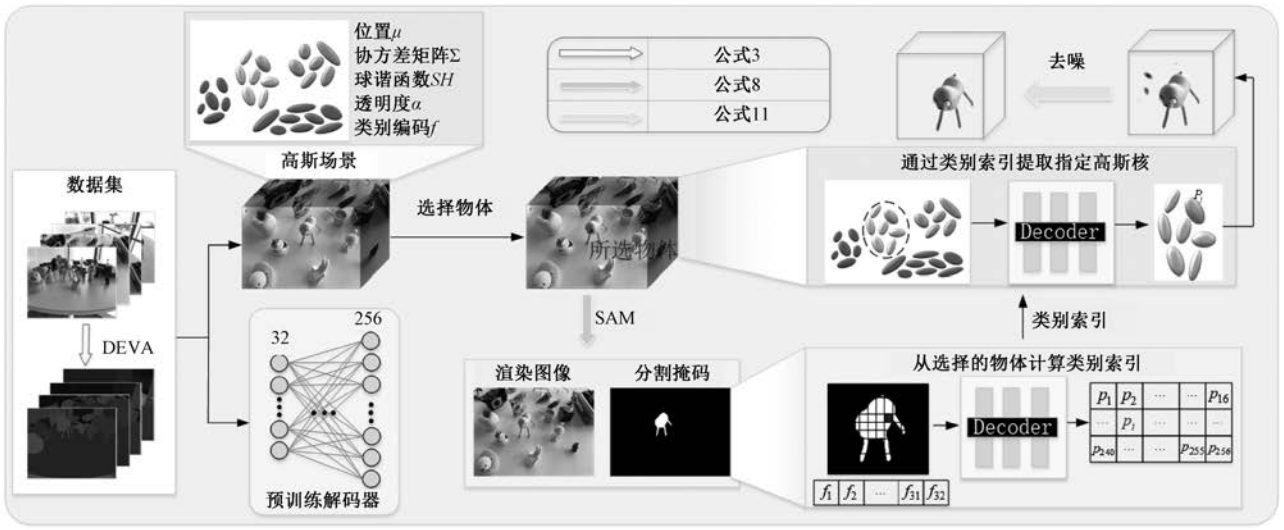

图 1 提出方法的整体管线

然后累加每个高斯核对其覆盖像素的颜色贡献，得到每个像素的颜色 C，其计算方式为

$$C=\sum_{i\in N}c_i\alpha_i\prod_{j=1}^{i-1}(1-\alpha_j) \tag{2}$$

其中，N 表示与像素重叠的有序高斯核集合，c_i 表示投影到当前像素的每个高斯核的颜色，α_i 表示高斯核在当前像素的透明度和权重的乘积。3DGS 提供了一种快速可微分的渲染器，可以实时渲染 3D 场景。

3.2 类别特征训练

为了在数据集中获得多视角一致的类别掩码，借助 DEVA[38] 模型对数据集进行连接和分割。该模型能够将连续图像序列中的相同对象分割为相同的掩码，在不同图像中，相同对象的灰度值保持一致。使用 DEVA 模型进行数据集预处理能确保获取多个视角一致的分割结果。这些灰度图用于监督训练。对于输入的图像数据集 I，相应的二维类别掩码可以表示为

$$I_i=DEVA(I) \tag{3}$$

其中，I_i 表示数据集 I 中每张图像的灰度分割结果，像素的每个值可以表示其像素所属物体的类别。多视角一致性的分割结果如图 2 所示。

图 2 灰度类别信息生成

如果想要区分重建场景中的不同对象，需要将图像中的类别信息训练到 3D 高斯核上，但是直接在高斯核上训练灰度值表示类别的分割结果是非常不准确的。为了在高斯核上保存更多的特征信息，本方法在每个高斯核上训练长度为 32 的类别特征，之后使用解码器进行类别预测。类别特征的训练过程借助 3DGS 的可微分渲染管线，与颜色计算过程类似，首先计算每个高斯核对图像像素的影响权重，随后，累加映射到每个像素的 N 个有序高斯核的类别特征，生成每个像素的特征。

$$F_p=\sum_{i\in N}f_i\alpha_i\prod_{j=1}^{i-1}(1-\alpha_j) \tag{4}$$

其中，F_p 表示由高斯核映射到图像每个像素的类别特征，f_i 表示每个高斯核上的类别特征，α_i 表示高斯核在当前像素的透明度和权重的乘积。

在本文方法中，解码器和类别特征是同时训练的，解码器用于将长度为 32 的类别特征解码为长度为 256 的概率分布，其中最大概率的索引表示与图像相对应的灰度值，解码器的结构如图 3 所示。

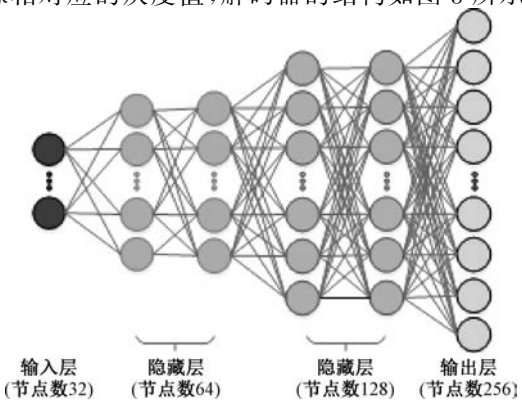

图 3 解码器的网络结构

在训练解码器 θ 时，假设图像 I 由 P 个像素组成，其中 F_p 代表每个像素的类别特征。因此，图像中每个像素的预测类别概率分布为

$$\tilde{y}_p = (\theta(F_p)) \tag{5}$$

在获得预测类别后，计算预测类别与真实图像类别之间的交叉熵损失，通过反向传播来逐渐更新解码器和高斯核上的类别特征，从而提高模型的分割的准确度。损失计算过程可以表示为

$$\mathcal{L}_c = -\frac{1}{N}\sum_{p=1}^{N} K_{N,p} \log \tilde{y}_{N,p} \tag{6}$$

其中，N 表示图像中的总像素数，$K_{N,p}$ 表示第 p 个像素的真实类别索引，$\tilde{y}_{N,p}$ 表示解码器对第 p 个像素的预测类别概率分布。为了训练重建场景，在渲染图像上使用 $L1$ 损失和结构性相似性损失 $L_{D\text{-}SSIM}$ 在模型训练中同时确保细节保真度和全局结构的一致性。因此整体模型的总训练损失为

$$\mathcal{L} = \mathcal{L}_c + (1-\lambda)\,\mathcal{L}_1 + \lambda\,\mathcal{L}_{D-SSIM} \tag{7}$$

3.3　模型分割与去噪

在进行模型分割时，首先需要获取目标对象的类别索引，然后提取出对应类别的高斯核。由于3DGS的输出通常以渲染图像形式呈现，类别特征也会随渲染管线渲染到图像中。因此，可以先利用图像分割模型生成目标对象的图像掩码，并通过解码器将掩码中的类别特征解码为目标对象的类别索引。与图像中的类别特征类似，所有高斯核的类别信息也以特征的形式存储，因此需要将所有高斯核的类别特征输入解码器，以获取每个高斯核的类别索引，从而实现目标模型的分割。

在模型分割时可以初步利用 SAM 图像分割模型或其他的图像分割模型通过输入提示 $\mathcal{P}$ 自动地分割渲染图像 I 。

$$M = SegModel(I, \mathcal{P}) \tag{8}$$

其中，M 表示图像分割模型输入分割提示后获得的分割掩码。随后，利用解码器 θ ，获得掩码内所有像素的类别概率分布。通过计算掩码区域内每个类别出现的概率，从而确定分割对象的类别索引，这一过程可以表示为

$$\hat{k} = \underset{g}{\operatorname{argmax}} \frac{1}{p}\sum_{i=1}^{\mathrm{p}} \delta(\theta(F_p), g) \tag{9}$$

其中，δ 表示克罗内克 δ 函数。随后，使用解码器获取所有高斯核的类别索引，找到对应的高斯核，从而实现 3D 模型的分割。整体的分割方法能有效地处理复杂场景和遮挡情况，提高分割的准确性和鲁棒性。

由于场景的高复杂性，直接提取对应类别的高斯核中会包含一些由离群点产生噪声。为解决这一问题，在高斯分割的过程中引入 KNN 算法。KNN 算法是一种常用的监督学习方法，广泛应用于分类和回归任务。在本方法中，通过比较每个高斯核与其 K 个最近邻高斯核的平均距离，判断该高斯核是否属于异常数据，平均距离的计算方式为

$$D_i = \frac{1}{K}\sum_{j=1}^{K} d_{ij} \tag{10}$$

其中，d_{ij} 表示第 i 个高斯与第 j 个最近邻高斯之间的距离。最后，通过距离阈值 D 去除噪声高斯：

$$\{P_i \mid D_i \geqslant D\} \tag{11}$$

4　实验结果与分析

4.1　实验细节设置

本文方法使用多种数据集验证方法的有效性，包括 Mip-NeRF360、LERF、LLFF 等开源数据集，并从数据集中选取不同场景进行实验分析，以上数据集内含有包含许多复杂对象的场景以及大型场景。本文的解码器基于 MLP，输入维度为 32，表示类别特征。输出维度为 256，表示预测类别的概率分布。在训练解码器时使用 Adam 优化方法进行优化，学习率设置为 $5e^{-4}$，迭代次数为 30000。KNN 算法去噪的距离阈值一般设置为整体空间大小的 1/100，最邻近粒子个数设置为 10。实验在 Pytorch2.0.1 框架下完成，实验平台为 Intel(R) Xeon(R) Gold 6133 CPU，NVIDIA GeForce RTX 4090 GPU。

4.2　定量分析

本文首先对训练过程的开销和分割质量进行了全面的定量分析。实验记录了训练过程中各种方法的内存消耗峰值，并通过评估性能指标证明本文方法在各种场景中的优越性。

4.2.1　训练开销

针对现有的分割方法所需的训练内存和训练时间太过昂贵的问题，本文对不同方法在不同数据集中的多种场景的训练内存和训练时间进行了详细分析，探索了多种技术以最大限度地减少内存需求，提升处理效率。为了降低内存开销并准确地分割出场景中的指定物体，本方法采用了长度为 32 的类别特征表示高斯核的类别信息，随后设计了一个简单高效的 MLP 解码器用于类别预测。实验对不同方法在不同数据集下的整体训练时间和训练过程中的内存占用峰值进行了定量比较，结果如表 1 所示。

表 1 不同方法的训练开销

	Gaussian Grouping[15]		SAGA[16]		Feature-3DGS[17]		本文方法	
	内存占用	训练时间	内存占用	训练时间	内存占用	训练时间	内存占用	训练时间
Figurines	38.35GB	58min	36.59GB	128min	17.8GB	18h	10.6GB	86min
Kitchen	22.5GB	53min	17.95GB	112min	17.4GB	18h	7.49GB	69min
Garden	41.90GB	50min	41.42GB	105min	17.5GB	18h	13.7GB	97min
Fortress	10.2GB	41min	10.6GB	93min	14.1GB	18h	8.3GB	65min
Bear	37.03GB	70min	20.1GB	68min	17.7GB	17h	9.2GB	83min
Horns	18.63GB	43min	14.5GB	79min	14.1GB	17h	9.3GB	81min
Ramen	18.02GB	43min	17.4GB	77min	14.1GB	17h	8.9GB	89min
Teatime	39.87GB	62min	23.1GB	64min	14.1GB	17h	9.3GB	77min

此外，实验对比了 SAGA[16] 和 Feature-3DGS[17] 使用文本提示进行分割的实验效果，实验结果如图 4 所示，在实验中，将分割出的三维模型所在位置高亮展示到原始图像上，以展示分割出的三维模型的准确性。其中，SAGA 的部分分割结果会产生一定程度的噪点，降低了分割算法的稳定性。另外，由于 Feature-3DGS 在训练过程中依赖于训练数据集中的文本标签进行监督学习，因此只能根据数据集中包含的标签进行分割，不支持开放性词汇，所支持的物体十分有限，实验表明，本文方法在节省大量内存开销和训练时间的同时，实现了更高的分割效果。

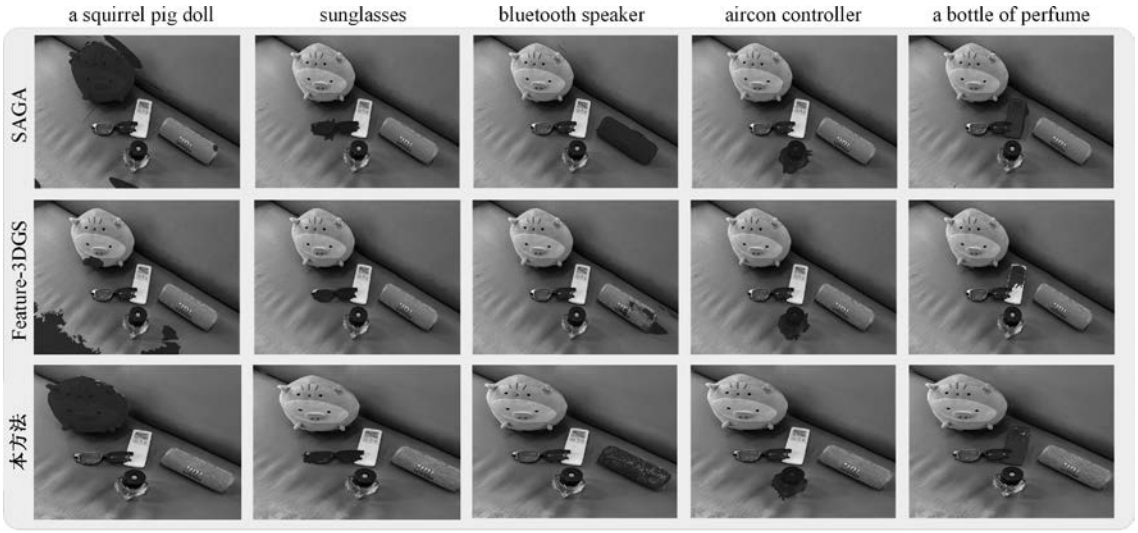

图 4 不同方法的分割效果

4.2.2 分割质量

为了衡量分割质量，采用了 mIoU 和 mBIoU 作为衡量指标，对各个方法的分割结果进行了计算。其中，mIoU 通过计算预测分割区域与实际标签区域的交集与并集的比值来衡量分类的精度，mBIoU 计算的是分割物体边界与真实物体边界的交并比，以衡量边界的准确性。实验选择了包含较多物体的 figurines 和 teatime 复杂场景数据集，计算所有物体分割质量的平均值，结果详见表 2 实验数据显示。本方法提出的特征表示方法与简洁高效的 MLP 解码器相结合，再使用 KNN 算法进行去噪，有效地提升了分割质量。这表明，在处理复杂场景时，该方法具备更好的适应性和有效性。通过对比各个方法的性能，可以看出，优化后的设计在资源利用率和分割精度方面均取得了良好的平衡。

表 2 不同方法的分割质量

场景方法	figurines		teatime	
	mIoU	mBIoU	mIoU	mBIoU
DEVA[38]	46.2	45.1	54.3	52.2
LERF[31]	33.5	30.6	49.7	42.6
SA3D[30]	24.9	23.8	42.5	39.2
LangSplat[36]	52.8	50.5	69.5	65.6
GaussianGrouping[15]	69.7	67.9	71.7	66.1
本文方法	89.2	84.4	85.7	79.7

4.3 定性分析

模型分割是在渲染图像上进行操作的。本文方法首先将类别特征映射到每个像素上，使用图像分割模型在图像上分割指定物体，再获取到分割部分对应像素的类别特征，将其输入解码器即可获得所

选物体的类别。之后,将高斯核上的类别特征同样放入解码器,解码出高斯核的类别,即可实现模型分割。因此,本文方法可以轻松地与现有的基础图像分割模型结合,只需要使用图像分割模型分割图像,获取所选物体对应的部分,就可以实现模型分割。本节展示了本文方法结合多种图像分割模型的分割结果。其中,单点提示、多点提示分割结合基础的SAM[9]模型,文本提示分割结合基于 Ground DINO[39]的 Lang-SAM 开源模型。实验证明了该方法在不同复杂场景中分割指定物体的能力,展现了本文方法的实用性和鲁棒性。

4.3.1　单点分割

使用单点输入作为提示的实验结果如图 5 和图 6 所示。该方法在各种场景中展现出稳定的性能,包括复杂背景和多个对象的场景。单点提示分割方法在速度上表现十分出色,分割时间保持在毫秒级范围内,极大提升了交互性。该方法的分割质量可与最先进的 NeRF 分割方法和基于 3DGS 的分割方法相媲美,且实现这种分割质量不需要大量的训练时间。与之前的基于 NeRF 的方法相比,所提方法提供了更精细的分割细节。此外,与类似工作 SAGA[16]相比,该方法实现更为简单,不仅能够达到相似的分割速度,还大大简化了实现过程,省去了复杂的预处理和训练步骤。

图 5　单点提示分割结果

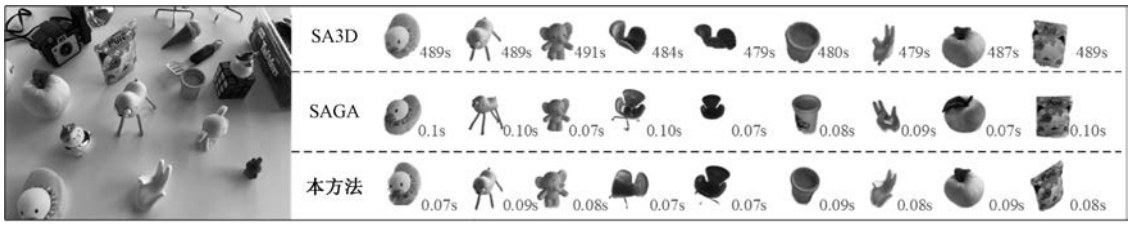

图 6　不同方法的分割时间

4.3.2　多点分割

本文方法还支持使用多点作为提示进行分割,分割结果如图 7 所示。该方法具有高度的通用性,能够通过点击多个物体作为输入提示,实现单个对象和多个对象的同时分割。这一能力在包含大量对象的复杂场景中尤为有利。通过利用多点提示,该方法增强了场景分割的鲁棒性和交互性,确保在存在重叠或密集排列对象的情况下也能获得准确的结果。与单点方法相比,多点提示作为分割条件提供了更高的多样性和准确性,使得该技术适用于更广泛的应用场景。

图 7　多点提示分割结果

本文在 LERF 数据集中的餐厅场景中,使用了一些文本提示词来引导分割。实验结果如图 8 所示。该方法能够在复杂的 3D 场景中通过文本提示准确识别指定对象,使得与 3D 模型的交互更加直观和灵活。该方法增强了 3D 分割在各种复杂场景中的可用性和适用性,为基于文本提示的精准模型提取提供了一种强大的解决方案。

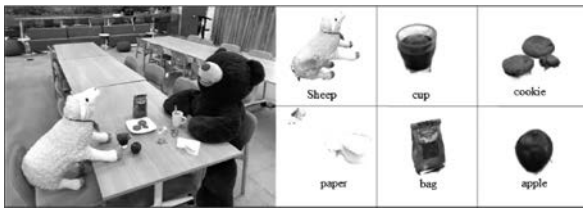

图 8　文本提示分割结果

4.3.3　任意角度分割

通过从当前相机位置渲染显示的图像,并利用各种提示对场景进行分割,该方法能够交互式地使用各种提示从不同视角的复杂场景中提取指定模型。如图 9 所示,实验结果展示了该方法在任意视角下对多种 3D 场景进行分割的结果。各个视角的分割结果清晰且细致,充分反映了模型对复杂背景和多个对象的处理能力。特别是在存在重叠和密集排列对象的情况下,该方法依然能够准确识别和提取目标,提升了操作的灵活性和效率。

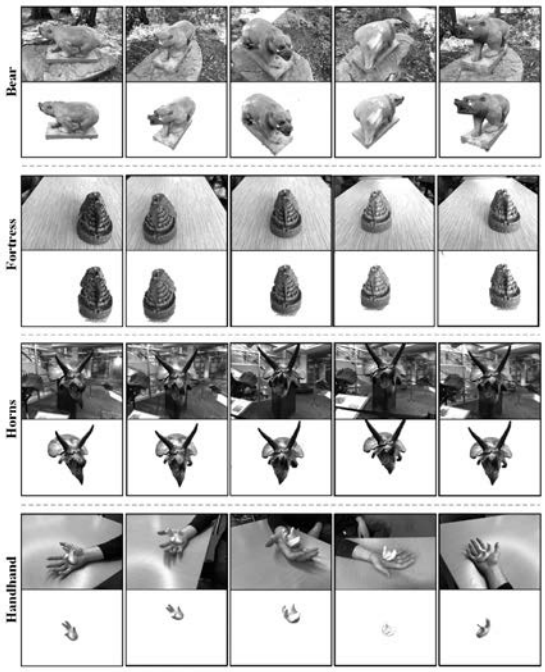

图 9 任意视角分割结果

4.3.4 去噪前后对比

在实验过程中，由于场景的复杂性和边界模糊问题，直接提取目标模型会产生一定的噪声，为了解决这个问题，采用了 KNN 算法将离群点视为异常点并去除。这种去噪算法优化了分割结果，提高分割过程的准确性。实验结果如图 10 所示，通过对比去噪前后的结果，展示了该方法的有效性，确保了更加准确的分割结果。

图 10 去噪前后分割结果对比

4.4 消融实验

4.4.1 不同特征长度

在本研究中，探索了使用不同长度的类别特征对高斯核进行解码。最初实验了长度为 16 的特征，但观察到这种特征长度结合简化的损失函数，导致分割模型有时受到不可修复的噪声影响。有限的特征长度不足以捕捉必要的细节，从而导致分割质量不佳。为了解决这一问题，增加了特征长度，以保留更多细节和更精确的特征。通过扩展特征长度，旨在提高解码器准确区分不同类别的能力。大量实验表明，长度为 32 的特征在稳定性和分割质量之间达到了最佳平衡，同时保持了合理的内存消耗，并提高了分割过程的整体鲁棒性。不同长度的实验结果如图 11 所示。

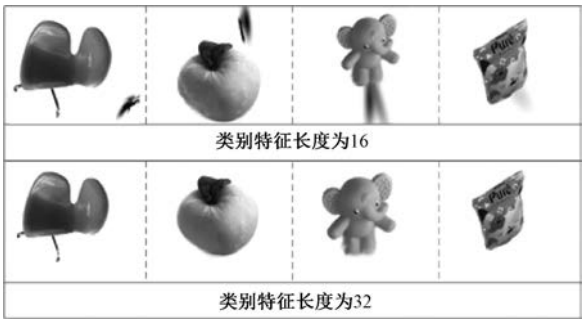

图 11 不同类别特征长度的分割结果

4.4.2 重建质量

最后，本文评估了论文中一些场景的重建质量和渲染效率，实验结果如表 3 所示，实验展示了引入类别编码对 3DGS 重建质量的影响。由于类别特征与颜色特征独立，因此该方法对原始 3DGS 渲染质量的影响可以忽略不计，由于额外需要渲染类别特征，会对渲染速度产生一定的影响，但是仍可以满足实时性的要求。本方法有效地增强了模型的特征表示能力，进一步分析表明，类别特征的引入不仅提升了模型对复杂场景中对象的识别能力，还提高了重建过程中对细节的捕捉能力。这种改进使得重建结果在保留对象形状和空间关系的同时，避免了由颜色特征引起的噪声和模糊。

表 3 添加类别特征的重建质量前后对比

	figurines		garden		teatime		fortress		kitchen		bear	
	3DGS	本方法	3DGS	本方法	3DGS	本方法	3DGS	本方法	3DGS	本方法	3DGS	本方法
PSNR	27.61	27.46	31.67	31.57	30.09	29.96	35.45	35.47	34.71	34.80	29.20	29.19
SSIM	0.9042	0.9031	0.9307	0.9297	0.9204	0.9198	0.9629	0.9629	0.9661	0.9661	0.9308	0.9311
LPIPS	0.1668	0.1681	0.0846	0.0886	0.1857	0.1870	0.0601	0.0603	0.0528	0.0529	0.1095	0.1095
FPS(帧/秒)	283	237	176	146	251	211	278	227	309	276	214	174

5 结　论

本文提出了一种新颖且轻量级的 3D 模型分割方法，构建一个简洁的分割模型，无需大量内存即可实现 3DGS 对重建场景中任意对象的分割。该方法能通过各种输入提示对三维模型进行分割，并且能够与已训练好的图像分割模型结合。此外，采用 KNN 算法进行去噪，该方法能去除模型分割时出现的噪声问题，增加分割结果的准确性。大量实验证明了本方法的效率和有效性，强调了其作为重建三维模型分割工具的潜力。

致　谢　感谢指导老师的细心指导，感谢评审专家的审稿意见和提出的专业性建议！

参　考　文　献

[1] Wang W, Neumann U. Depth-aware cnn for rgb-d segmentation//Proceedings of the European Conference on Computer Vision. Munich, Germany,2018: 135-150

[2] Hou J, Dai A, Nießner M. 3d-sis: 3d semantic instance segmentation of rgb-d scans//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. California, USA, 2019: 4421-4430

[3] Qi C R, Su H, Mo K, et al. Pointnet: deep learning on point sets for 3d classification and segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 652-660

[4] Qi C R, Yi L, Su H, et al. Pointnet++: deep hierarchical feature learning on point sets in a metric space//Proceedings of the 31st International Conference on Neural Information Processing Systems. California, USA, 2017: 5105-5114

[5] Huang J, You S. Point cloud labeling using 3d convolutional neural network//2016 23rd International Conference on Pattern Recognition. Cancun, Mexico, 2016: 2670-2675

[6] Tang H, Liu Z, Zhao S, et al. Searching efficient 3d architectures with sparse point-voxel convolution//Proceedings of the European Conference on Computer Vision. Glasgow, UK, 2020: 685-702

[7] Ye D, Zhou Z, Chen W, et al. Lidarmultinet: towards a unified multi-task network for lidar perception//Proceedings of the AAAI Conference on Artificial Intelligence. Washington, USA,2023: 3231-3240

[8] Zhou H, Zhu X, Song X, et al. Cylinder3d: an effective 3d framework for driving-scene lidar semantic segmentation, https://doi.org/10.48550/arXiv.2008.01550 2020,8,4

[9] Kirillov A, Mintun E, Ravi N, et al. Segment anything//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France,2023: 29914-29934

[10] Ke L, Ye M, Danelljan M, et al. Segment anything in high quality//Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA, 2023, 36

[11] Peng S, Genova K, Jiang C, et al. Openscene: 3d scene understanding with open vocabularies//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada,2023: 815-824

[12] Zhang J, Dong R, Ma K. Clip-fo3d: learning free open-world 3d scene representations from 2d dense clip//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France, 2023: 2048-2059

[13] Mildenhall B, Srinivasan P P, Tancik M, et al. Nerf: representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 2021, 65(1): 99-106

[14] Kerbl B, Kopanas G, Leimkühler T, et al. 3D gaussiansplatting for real-time radiance field rendering. ACM Transactions on Graph., 2023, 42(4): 139:1-139:14

[15] Ye M, Danelljan M, Yu F, et al. Gaussian grouping: segment and edit anything in 3d scenes//Proceedings of the European Conference on Computer Vision. Milano,Italy, 2024: 162-179

[16] Chen J Z, Fang J M, Yang C, et al. Segmentany3d gaussians, https://doi.org/10.48550/arXiv.2312.00860 2023, 12,1

[17] Zhou S, Chang H, Jiang S, et al. Feature 3dgs: supercharging 3d gaussian splatting to enable distilled feature fields//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2024: 21676-21685

[18] Barron J T, Mildenhall B, Tancik M, et al. Mip-nerf: amultiscale representation for anti-aliasing neural radiance fields//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal,Canada, 2021: 5855-5864

[19] Barron J T, Mildenhall B, Verbin D, et al. Mip-nerf 360: unbounded anti-aliased neural radiance fields//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA, 2022: 5470-5479

[20] Barron J T, Mildenhall B, Verbin D, et al. Zip-nerf: anti-aliased grid-based neural radiance fields//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France, 2023: 19697-19705

[21] Chen A, Xu Z, Zhao F, et al. Mvsnerf: fast generalizable radiance field reconstruction from multi-view stereo//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada,2021: 14124-14133

[22] Yu A, Ye V, Tancik M, et al. Pixelnerf: neural radiance fields from one or few images//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA,2021: 4578-4587

[23] Luiten J, Kopanas G, Leibe B, et al. Dynamic 3d gaussians:

tracking by persistent dynamic view synthesis//Proceedings of the International Conference on 3D Vision 2024. Davos, Switzerland, 2024: 800-809

[24] Wu G, Yi T, Fang J, et al. 4d gaussian splatting for real-time dynamic scene rendering//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2024: 20310-20320

[25] Tang J, Ren J, Zhou H, et al. Dreamgaussian: generative gaussian splatting for efficient 3d content creation, https://arxiv.org/abs/2309.16653 2023,9,28

[26] Yi T, Fang J, Wu G, et al. Gaussiandreamer: fast generation from text to 3d gaussian splatting with point cloud priors//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2024: 6796-6807

[27] Rombach R, Blattmann A, Lorenz D, et al. High-resolution image synthesis with latent diffusion models//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA,2022: 10684-10695

[28] Zhi S, Laidlow T, Leutenegger S, et al. In-place scene labelling and understanding with implicit scene representation//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal,Canada,2021: 15838-15847

[29] Ren Z, Agarwala A, Russell B, et al. Neural volumetric object selection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA,2022: 6133-6142

[30] Cen J, Zhou Z, Fang J, et al. Segment anything in 3d with nerfs. Advances in Neural Information Processing Systems, 2023, 36: 25971-25990

[31] Kerr J, Kim C M, Goldberg K, et al. Lerf: language embedded radiance fields//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France,2023: 19729-19739

[32] Goel R, Sirikonda D, Saini S, et al. Interactive segmentation of radiance fields//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada,2023: 4201-4211

[33] Wang B, Chen L, Yang B. Dm-nerf: 3d scene geometry decomposition and manipulation from 2d images, https://doi.org/10.48550/arXiv.2208.07227 2022,8,15

[34] Mirzaei A, Aumentado-Armstrong T, Derpanis K G, et al. SPIn-NeRF: Multiview segmentation and perceptual inpainting with neural radiance fields//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada,2023: 20669-20679

[35] Yin Y, Fu Z, Yang F, et al. Or-nerf: Object removing from 3d scenes guided by multiview segmentation with neural radiance fields, https://doi.org/10.48550/arXiv.2305.10503 2023,5,17

[36] Qin M, Li W, Zhou J, et al. Langsplat: 3d language gaussian splatting//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2024: 20051-20060

[37] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision//Proceedings of the International Conference on Machine Learning. Vienna, Austria, 2021: 8748-8763

[38] Cheng H K, Oh S W, Price B, et al. Tracking anything with decoupled video segmentation//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France,2023: 1316-1326

[39] Liu S, Zeng Z, Ren T, et al. Grounding dino: marrying dino with grounded pre-training for open-set object detection//Proceedings of the European Conference on Computer Vision. Milano,Italy,2024: 38-55

WANG Feng, Ph. D. candidate. His research interests include computer graphics, computer vision, 3D reconstruction and physical simulation.

YIN Ying, M. S. candidate. Her research interests include computer vision and 3D reconstruction.

WANG Jia-Yan, Ph. D. candidate. Her research interests include computer image processing and digital watermarking.

TANG Yong, Ph. D., professor. His research interests include virtual reality technology and application.

LI Sheng, Ph. D., professor. His research interests include computer graphics and virtual simulation technology.

ZHAO Jing, Ph. D., associate professor. Her research interests include virtual reality technology and application, computer vision, physical simulation.

Background

This work introduces a novel and lightweight pipeline for 3D model segmentation based on the recently developed 3D Gaussian Splatting (3DGS). As the demand for high-quality 3D visualizations grows across various industries, including virtual reality, gaming, and medical imaging, the need for efficient and accurate segmentation methods has become increasingly critical. Our method effectively integrates with existing 2D segmentation models, enhancing the accuracy and efficiency of object segmentation in complex 3D scenes.

3D Gaussian Splatting has shown exceptional promise in surpassing traditional approaches like Neural Radiance Field (NeRF) in both training speed and reconstruction quality. 3DGS enables the representation of 3D objects in a way that captures intricate details while maintaining rapid computa-

tion, making it a compelling tool for 3D visualization and editing. However, despite these advantages, segmentation methods tailored for 3DGS remain underdeveloped, which limits its application in detailed scene analysis and editing.

Our proposed pipeline addresses this gap by introducing an interactive segmentation method that not only improves segmentation quality but also operates with minimal memory overhead. The approach allows users to segment objects in 3D environments efficiently, leveraging the unique capabilities of 3DGS to handle complex geometries and lighting conditions effectively. By integrating user input in the form of point clicks or text prompts, the segmentation process becomes more intuitive and accessible, catering to a broader range of applications.

In summary, this work presents a significant advancement in the field of 3D model segmentation by combining the strengths of 3D Gaussian Splatting with innovative interactive techniques. The proposed pipeline not only enhances segmentation quality but also ensures efficient memory usage, making it suitable for real-time applications. By addressing the limitations of current segmentation methods for 3DGS, this research opens new avenues for detailed scene analysis and interactive editing, paving the way for improved experiences in virtual reality, gaming, and medical imaging. Through rigorous experimentation and validation, the effectiveness of the proposed methods is demonstrated, providing a robust solution for the challenges faced in the segmentation of 3D models.

This work was supported by grants from the Science and Technology Program of Hebei(No. 246Z0105G) and Innovation Capability Improvement Plan Project of Hebei Province (No. 22567626H).